

Editorial

Modelos Psicométricos e seu Impacto na Justiça na Avaliação

Víthor Rosa Franco

Universidade São Francisco – USF, Campinas-SP, Brasil

Neste editorial, discutirei a relação entre métodos psicométricos e como o uso desses métodos está atrelado à justiça na avaliação. Os leitores deste periódico certamente estão acostumados com métodos tradicionais da psicometria, como a Análise Fatorial e a Teoria de Resposta ao Item (TRI), ou mesmo com modelos um pouco mais modernos, mas que ainda fazem parte do *mainstream* da psicometria, como os Modelos Diagnósticos Cognitivos (CDMs) e as análises de rede. Esses métodos são ferramentas poderosas que permitem mensurar habilidades e traços psicológicos de maneira mais robusta, oferecendo também informações sobre as propriedades dos instrumentos de avaliação. No entanto, é essencial conhecer as limitações de cada método e, acima disso, entender que cada uma dessas ferramentas carrega uma série de pressupostos teóricos que pode comprometer a justiça na avaliação. Nesse sentido, este editorial reforça que "para quem só sabe usar um martelo, tudo é prego", nos lembrando que a dependência exclusiva de um método pode levar a simplificações ou distorções.

É necessário entender que quando falamos em métodos psicométricos estamos falando em modelos de medida que são utilizados para formalizar as relações entre construtos psicológicos e os comportamentos das pessoas (Baptista & Franco, 2023). Podemos dissecar essa definição em cada uma de suas partes fundamentais para então começar a entender as limitações mais implícitas dos métodos psicométricos. Assim, em primeiro lugar, métodos psicométricos são “modelos”, o que significa que as análises psicométricas, além de poderem ser classificadas como métodos estatísticos, também refletem alguma teoria psicológica. Esse fato é muitas vezes deixado de lado e, por isso, muitas pessoas cometem (tomam?) uma série de decisões sobre a construção de instrumentos que levam a práticas inadequadas de mensuração (Flake & Fried, 2020). Em seguida, precisamos compreender como os modelos formalizam as relações entre construtos psicológicos e os comportamentos observados.

Franco et al. (2022) qualificam essa discussão enfatizando que há ao menos três tipos de pressupostos implícitos inerentes aos modelos psicométricos tradicionais. O primeiro pressuposto, chamado de validade estrutural, está relacionado às escolhas mais operacionais dos modelos matemáticos subjacentes à análise. Por exemplo, no contexto da análise de itens dicotômicos usando métodos da TRI, existe a família de modelos logísticos, sendo comumente utilizados os modelos de um, dois, ou três parâmetros. Esse tipo de pressuposto é chamado de “validade estrutural” pois ele indica que se assume que a parametrização escolhida para realizar a análise de um teste é a mais adequada. Apesar de a prática de comparação de índices de ajuste ajude a minimizar os impactos desse pressuposto, não é possível comparar um modelo que não foi testado. Portanto, corre-se sempre o risco de que entre os modelos escolhidos para comparação não esteja, de fato, o modelo mais adequado.

O segundo pressuposto implícito identificado por Franco et al. (2022) é o pressuposto de processo, o qual se refere às crenças sobre o processo subjacente que gera os dados observados. Tal pressuposto é identificado na literatura (van Bork et al., 2024) como um dos motivos para haver equivalências entre métodos da teoria clássica dos testes, da análise fatorial e da TRI. De fato, apesar de operacionalmente esses métodos serem bastante distintos, todos eles suportam o mesmo pressuposto de que há uma relação monotônica entre o traço latente e maiores escores brutos em um teste. Métodos alternativos como os modelos de desdobramento (Balafas et al., 2020) e os modelos de árvores de processamento multinomial (ou, mais comumente na literatura psicométrica, os IRTree; Xue & Chen, 2024) oferecem uma visão diferente sobre como a variável latente impacta nas respostas observadas.

Por fim, Franco et al. (2022) discutem o pressuposto de construto: a crença de que os dados observados, por si só, não constituem uma medida, sendo o objetivo da psicometria a identificação da magnitude ou classe de variáveis latentes que originam os dados observados. Esse pressuposto é provavelmente o mais central a toda a psicometria, sendo

que a maioria dos modelos utilizados na área podem ser categorizados sob o guarda-chuva da abordagem dos Modelos Generalizados de Variáveis Latentes (Noventa et al., 2024). Mais especificamente, esses modelos assumem que as variáveis latentes são a causa dos comportamentos observados, sendo conhecidos também como modelos reflexivos (Bollen & Ting, 2000). Sob esse guarda-chuva estão todos os modelos de análise fatorial, TRI paramétrica, não-paramétrica, de mistura e multidimensional, modelos de desdobramento, modelos de árvores de processamento, CDMs, análise de perfis e de classes latentes, entre outros (Mair, 2018).

Neste ponto, explicita-se os dois problemas principais subjacentes que definem as limitações geradas por esses pressupostos. O primeiro problema é que modelos que não são testados, nunca serão identificados como os mais apropriados aos dados. O segundo problema é que há modelos que podem parecer distintos, mas que são matematicamente equivalentes. É importante notar que esses problemas vão além da prática inadequada da psicometria, como definida por Flake e Fried (2020), mas envolvem uma compreensão mais aprofundada de o que de fato as técnicas de análise de dados em psicometria nos ajudam a “revelar” sobre os comportamentos das pessoas. Além disso, a reflexão sobre tais problemas nos ajuda a entender como as análises quantitativas, apesar de serem fundamentadas na matemática, muitas vezes não são tão objetivas como se vende (Jamieson et al., 2023).

Sobre o problema da equivalência entre modelos, a princípio, existe uma solução que pode ser vista como trivial: basta que todos nós, psicólogos, estudemos mais matemática para aprendermos como identificar esse problema. Apesar de tal solução poder ser aversiva a como os currículos em psicologia têm se estruturado ao longo do tempo (Franco & Iglesias, 2023), tem se defendido tal solução como aspecto importante à redução de práticas questionáveis de pesquisa na área (Flake & Fried, 2020). De fato, a psicologia se constitui, quando se avaliam as práticas hegemônicas de pesquisa, principalmente como uma ciência quantitativa, seja isso positivo ou não) (Nuttgens, 2023) e, portanto, é razoável esperar que pesquisadores da área tenham um maior conhecimento sobre análise de dados (em um nível mais operacional) e matemática (em um nível teórico).

No entanto, essa solução é inviável em um grau importante. A psicologia, obviamente, não está confinada apenas ao contexto de pesquisa e ter a formação desses profissionais voltada exclusivamente para a pesquisa significaria sacrificar outros aspectos fundamentais relacionadas à prática da profissão (de Sousa & Lima, 2024). Mesmo que o raciocínio científico tenha seu valor para a prática na área (no Brasil, institucionalizado pelo código de ética da profissão; Conselho Federal de Psicologia, 2005), a formação desses profissionais deve considerar escopos mais amplos da compreensão dos fenômenos psicossociais, o que envolve também aspectos culturais, sociais e políticos que vão além da formalidade das teorias científicas da área. Assim, a “solução” apontada anteriormente pode ter uma consequência positiva para as práticas de pesquisa na área, mas colocam em risco o desenvolvimento de outros aspectos fundamentais da atuação profissional.

Em nível mais individual, nada impede que pesquisadores da psicometria busquem se aprofundar nessa temática. Por exemplo, Kruis e Maris (2016) apresentam um estudo matemático aprofundado das situações nas quais modelos reflexivos, formativos e de redes serão equivalentes. Enquanto modelos reflexivos assumem que variáveis latentes causam o comportamento observado, e os modelos de rede que os comportamentos se causam entre si, modelos formativos assumem que as representações que criamos de construtos são apenas a soma dos comportamentos observados. De fato, há uma discussão extensa na literatura (Baptista & Franco, 2024; Bollen & Diamantopoulos, 2017; Bollen & Ting, 2000; Cunha et al., 2023; Hallquist et al., 2020; Kan et al., 2020; Molenaar, 2010; Ting, 1998) sobre como cada um desses modelos impacta na nossa interpretação dos processos psicológicos que estudamos, mas que, em muitos casos, não é possível levantar evidências robustas que nos permitam concluir qual desses modelos melhor explica o comportamento.

Essa questão nos leva, então, ao problema de que modelos que não são testados nunca serão identificados. Esse problema, em poucas palavras, não tem uma solução fácil. De fato, esse problema é o cerne do processo da descoberta científica, o qual envolve, justamente, a busca pelo modelo (ou teoria) que melhor explique os fenômenos de interesse (Stanovich, 2021). No entanto, no contexto da psicometria, é importante que tenhamos a noção de que as escolhas analíticas que fazemos determinam as conclusões que chegamos sobre o que constitui um “bom” instrumento de medida em psicologia. Nesse sentido, a prática hegemônica tem sido de aceitar um subconjunto de métodos de análise da abordagem dos Modelos Generalizados de Variáveis Latentes como a única possibilidade de criação de medidas em psicologia.

Finalmente, compreendidas as limitações das práticas convencionais da psicometria, podemos entender como elas impactam na justiça em avaliação: métodos psicométricos distintos nos levam a conclusões distintas. Apesar de parecer uma obviedade, esse fato é geralmente deixado de lado no desenvolvimento de instrumentos de avaliação. Podemos pensar, por exemplo, na distinção do uso da TRI ou da Teoria de Espaços do Conhecimento (*Knowledge Space Theory*, KST; Doignon & Falmagne, 1999) para definir o desempenho de estudantes. A TRI, sendo um modelo que assume que a variável latente é contínua, define o espaço de aprendizagem como um espaço linear: conforme se estuda mais sobre uma temática, há um crescimento em todas as habilidades relacionadas àquela temática. Na KST, por outro lado, assume-se que as pessoas têm estados distintos de conhecimento e que a aprendizagem não precisa ser linear. Nesse

sentido, estudar mais sobre um tema permite às pessoas desenvolverem novas habilidades, sendo que cada pessoa pode seguir um caminho distinto de aprendizagem e, portanto, nem todos os conhecimentos são cumulativos.

Essas diferenças teóricas, que nesse caso resultam também em algumas diferenças analíticas, podem impactar, e muito, processos sociais dependentes de avaliações. Por exemplo, Marsura (2023) investigou o impacto que há em utilizar a KST ao invés da TRI para determinar o nível de conhecimento de respondentes no ENEM. Os resultados mostraram que os métodos não convergem totalmente na identificação dos indivíduos com maior desempenho e, portanto, caso a KST fosse utilizada ao invés da TRI, pessoas diferentes poderiam ser aprovadas no exame. Como a escolha do modelo analítico é, em certa dimensão, arbitrária, temos como consequência também que a decisão de quem é ou não aprovado também é, em certa dimensão, arbitrária. No entanto, os impactos nas vidas das pessoas que não são ingressam no ensino superior por causa do desempenho no ENEM podem ser imensos.

Muitos podem fazer uma leitura niilista do último parágrafo de que a psicometria e a avaliação psicológica estariam destinadas ao fracasso para aplicações no mundo real. No entanto, a principal mensagem deste editorial é que a reflexividade (Jamieson et al., 2023) deve ser uma constante no processo de desenvolvimento de instrumentos psicométricos. Afinal, já sabemos da necessidade de investigar e prover evidências diversas de validade e de fidedignidade. Portanto, a “arbitrariedade” discutida no parágrafo anterior não é tão certa assim se tivermos evidências robustas de que as medidas geradas, de fato, refletem o processo que estamos avaliando. Para isso, é importante que não cedamos às práticas muitas vezes dogmáticas da área e que também busquemos conhecer o que mais pode ser a psicometria, para além de seus modelos e pressupostos convencionais.

Referências

- Balafas, S. E., Krijnen, W. P., Post, W. J., & Wit, E. C. (2020). Mudfold: An R package for nonparametric IRT modelling of unfolding processes. *The R Journal*, 12(1), 49-75. <https://doi.org/10.32614/RJ-2020-002>
- Baptista, M. N., & Franco, V. R. (2024). Uma análise teórica da depressão sob a visão dos modelos psicométricos hegemônicos. *Estudos e Pesquisas em Psicologia*, 24. <https://doi.org/10.12957/epp.2024.67278>
- Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, 22(3), 581-596. <https://doi.org/10.1037/met0000056>
- Bollen, K. A., & Ting, K.-f. (2000). A tetrad test for causal indicators. *Psychological Methods*, 5(1), 3-22. <https://doi.org/10.1037/1082-989X.5.1.3>
- Conselho Federal de Psicologia (2005). *Código de ética profissional do psicólogo*. Conselho Federal de Psicologia.
- Cunha, R. D., Faiad, C., Baptista, M. N., & Cardoso, H. F. (2023). O mecanismo de geração de dados: relação entre construtos e seus indicadores. *Psicologia: Teoria e Pesquisa*, 39, e39nspe08. <https://doi.org/10.1590/0102.3772e39nspe08.en>
- de Sousa, L. T. M., & Lima, L. S. (2024). O histórico da constituição da psicologia enquanto ciência e profissão: por uma psicologia crítica latino-americana para a América Latina. *Revista Esperança Garcia*, 1(01). <https://revista.esperancagarcia.uespi.br/index.php/reg/article/view/10>
- Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. Springer.
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456-465. <https://doi.org/10.1177/2515245920952393>
- Franco, V. R., & Iglesias, F. (2023). Convite à Psicologia Matemática: Modelos e Benefícios da Teorização Formal. *Psicologia: Teoria e Pesquisa*, 39, e39515. <https://doi.org/10.1590/0102.3772e39515.pt>
- Franco, V. R., Laros, J. A., Wiberg, M., & Bastos, R. V. S. (2022). How to think straight about psychometrics: Improving measurement by identifying its assumptions. *Trends in Psychology*, 1-21. <https://doi.org/10.1007/s43076-022-00183-6>
- Hallquist, M. N., Wright, A. G., & Molenaar, P. C. (2021). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate Behavioral Research*, 56(2), 199-223. <https://doi.org/10.1080/00273171.2019.1640103>
- Jamieson, M. K., Govaart, G. H., & Pownall, M. (2023). Reflexivity in quantitative research: A rationale and beginner's guide. *Social and Personality Psychology Compass*, 17(4), e12735. <https://doi.org/10.1111/spc3.12735>
- Kan, K. J., de Jonge, H., van der Maas, H. L., Levine, S. Z., & Epskamp, S. (2020). How to compare psychometric factor and network models. *Journal of Intelligence*, 8(4), 35. <https://doi.org/10.3390/jintelligence8040035>
- Kruis, J., & Maris, G. (2016). Three representations of the Ising model. *Scientific Reports*, 6(1), 34175. <https://doi.org/10.1038/srep34175>
- Mair, P. (2018). *Modern psychometrics with R*. Springer.
- Marsura, A. C. (2023). *Avaliação das estruturas de conhecimento do Enem: Grupos diferentes aprendem diferentemente?* Dissertação de Mestrado, Programa de Pós-graduação em Psicologia, Universidade São Francisco, Campinas, São Paulo.
- Molenaar, P. C. M. (2010). Latent variable models are network models. *Behavioral and Brain Sciences*, 33(2-3), 166-166. <https://doi.org/10.1017/S0140525X10000798>
- Noventa, S., Heller, J., & Kelava, A. (2024). Toward a unified perspective on assessment models, part I: Foundations of a framework. *Journal of Mathematical Psychology*, 122, 102872. <https://doi.org/10.1016/j.jmp.2024.102872>
- Nuttgens, S. (2023). Making Psychology “Count”: On the Mathematization of Psychology. *Europe's Journal of Psychology*, 19(1), 100-112. <https://doi.org/10.5964/ejop.4065>
- Stanovich, K. E. (2021). *How to think straight about psychology*. HarperCollins Publishers.

-
- Ting, K. F. (1998). The TETRAD approach to model respecification. *Multivariate Behavioral Research*, 33(1), 157-154. https://doi.org/10.1207/s15327906mbr3301_7
- van Bork, R., Rhemtulla, M., Sijtsma, K., & Borsboom, D. (2024). A causal theory of error scores. *Psychological Methods*, 29(4), 807-826. <https://doi.org/10.1037/met0000521>
- Xue, M., & Chen, Y. (2024). A Stan tutorial on Bayesian IRTree models: Conventional models and explanatory extension. *Behavior Research Methods*, 56(3), 1817-1837. <https://doi.org/10.3758/s13428-023-02121-5>

Como citar este artigo

Franco, V. R. (2024). Editorial: Modelos Psicométricos e seu Impacto na Justiça na Avaliação. *Avaliação Psicológica*, 23(4), A-D. <http://dx.doi.org/10.15689/ap.2024.2304.ed>